ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 4, 374-380 (2020)

www.astesj.com

Special Issue on Innovative Research in Applied Science, Engineering and Technology

ASTESJ

ISSN: 2415-6698

Survey Analysis: Enhancing the Security of Vectorization by Using word2vec and CryptDB

Hana Yousuf[1], Said Salloum[1,2,*]

[1]*Faculty of Engineering &IT, The British University in Dubai, 345015, United Arab Emirates*

[2]*Research Institute of Sciences & Engineering, University of Sharjah, 27272, United Arab Emirates*

ARTICLE INFO

Article history:
Received: 24 June, 2020
Accepted: 17 July, 2020
Online: 28 July, 2020

Keywords:
Natural Language Processing
Vectorisation
Word2vec
CryptDB

ABSTRACT

Vectorization is extracting data from strings through Natural Language Processing by using different approaches; one of the best approaches used in vectorization is word2vec. To make the vectorized data secure, we must apply a security method, which will be CryptDB. The paper is analyzing the survey, which is created to interview security engineers through the SPSS tool. By analyzing the responses from software security engineers, it is seen that both word2vec and CryptDB works significantly. Word2vec is an effective vectorization approach, while CryptDB is an effective, secure database. In future work, we will be developing a secure vectorization using both approaches.

1. Introduction

A vector is a set of real numbers that is converted from a string. This is performed to extract useful data from a specific work using Natural Language Processing (NLP) techniques through deep learning and machine learning techniques [1]–[6]. This conversion is known as vectorization or word embedding. It converts multiple words or an entire phrase to the corresponding vector format of real numbers that are utilized to predict the words and similarities. This would be helpful for classifying texts, identifying similar words, clustering the documents, or extracting the features [7]. On the whole, vectorization is a term referring to the conversion of scalar instructions to vector instructions. The data is stored in vector format, where the loop is reconfigured in such a way that multiple elements can be processed in place of a single array. Vectorisation is used for executing systematic computation where large amounts of data must be processed effectively. Since the size of the code is reduced, the number of bugs is also reduced. It is also easier to substitute the values in mathematical expressions, thereby making the calculation easier and also easier to read and the data can be stored in compact manner.

Technically speaking, in word embeddings, the individual words are characterized as vectors with real value in a predefined cell [8]. The individual words are mapped to their respective vectors. The values of these vectors are trained similarly to neural networks; hence they are often grouped under deep learning techniques. Representations are provided based on the similarities between the words, thereby understanding their meanings. An embedding layer is a part of the work embedding which is trained jointly with neural network models on specific NLP like classifying the documents or language modelling [9]. The document containing the lexical data must be cleaned so that the individual words are hot encoded. They are initially initialized with smaller random numbers. The method of training embedding layers needs a large amount of training data. This approach of learning an embedding layer requires a lot of training data and can be slow, but will learn an embedding both targeted to the specific text data and the NLP task.

There are many word embedding approaches where the modes have been pre-trained. Some of them are word2vec, Glove, and Fast text from Google, Stanford, and Facebook, respectively [10]. Word2vec was developed by Toma Mikolov and other researchers from Google in 2013 [11]. It is necessary to understand why word2vec was created. The other NLP systems at that period used to consider the words as individual atomic units. There were no similarities between the words, and the models were no effective for smaller datasets. Hence, larger datasets with complicated models have used neural network approaches for training complicated data models and works better for larger datasets with lots of vocabulary. Therefore, since the neural network is used in word2vec, it works effectively in measuring the quality of the resulting vectorized contents. Hence, the degree of similarity is considered for identifying similar words and grouping them.

*Corresponding Author: Said Salloum, University of Sharjah, UAE. Tel: +971507679647 Email: ssalloum@sharjah.ac.ae

In order to improve the architecture for the word representations, the accuracy must be improved, and the computational complexity must be minimized. Hence, neural network models like Recurrent and Feedforward neural networks are linked to Net Language Model (NLM) [12]. They can be trained through Backpropagation or Stochastic gradient descent. In FFNNLM, it contains the layers like input, hidden, and output layers like any other work. In addition, it also contains the projection layer. Since the computation between hidden layers and projection layers are complicated, the data in the projection layer becomes denser [13]. In RNNLM, the model has more complicated patterns and layers; hence projection layer is not required in this model. Only the three basic layers of the deep network are sufficient.

There are two different models for learning that can be part of the word2vec technique for learning the word embeddings. They are:

- Continuous Bag-of-Words (CBOW)
- Continuous Skip-Gram

The former gets trained based on the prediction of the present word based on its context, while the latter gets trained by learning the surrounding words, which will lead to the present word. Both of them focus on learning the words concerning their context and usage and also the configurable parameters of the models. The major benefit of the technique is that higher-quality word embeddings may be trained effectively with reduced space and time requirements. This allows bigger embeddings to get trained with higher magnitudes and more lexical data. The architecture of the word2vec algorithm is given in Figure 1.

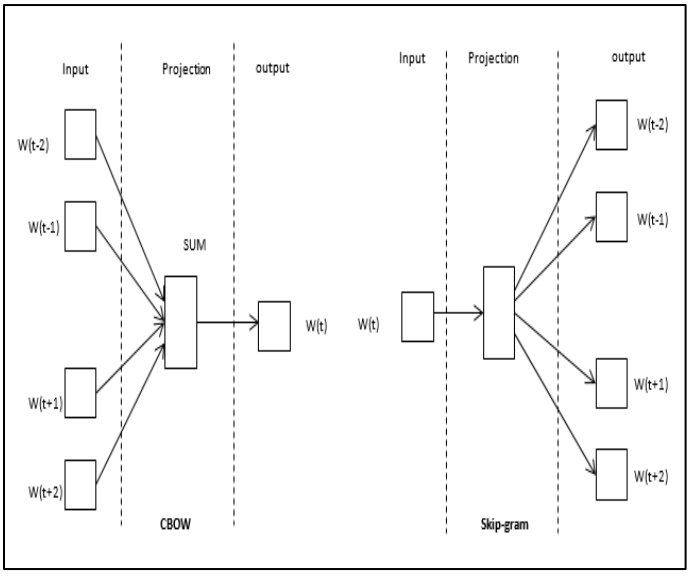

Figure 1: Architecture of word2vec

Network Functions Virtualisation (NFV) is a network architecture that utilizes IT virtualization to virtualize the entire classes of network node functions into structural blocks [14], [15]. Firewalls and network intrusions detections are necessary to provide security to the hardware. Hence these must be utilized in the application layer of the system [16]. Vectorisation is gradually taking center stage, which is being utilized in hardware equipment [17]. Newer processors are altering their architectures to improve the security for vectorization [18]. In these CPUs, it is obligatory to achieve vectorization for attaining higher performance [19]. When vectorized query execution is utilized, the processes are streamlined by processing multiple blocks at the same time. This may usually be 1024 blocks. Within the block, every column is saved as a single array, i.e., vector. Simple arithmetic operations and comparisons are performed by rapid iterations of vectors in a tight loop with very few function calls or conditional branches in the loop. These loops compile in a streamlined manner with fewer instructions and complete the instructions with a reduced clock cycle by utilizing the processors and cache memory.

The databases may be vulnerable to sensitive information or theft and can be attacked since the attackers can take advantage of the bugs in order to gain access to private information. Also, the attackers may seize and leak the data to those who require them. Hence, the confidentiality and piracy of the database must be safeguarded. CryptDB provides confidentiality against the attackers through SQL databases [20]. The query processes in the SQL are encrypted by collecting effective SQL aware encryption schemes. The process can also be used for encrypting the credential keys so that an item can be decrypted with merely a password. Since the processes and data are encrypted, the database administrator will not be able to view the data. In case of an attack on the server, the invader would not be able to gain access to the decrypted data. The database is capable of handling multiple queries simultaneously with fewer maintenance costs.

2. Literature Review

A study by [21] has assessed security risks through both quantitative and qualitative analysis. A critical review of the existing standards has been performed [22], [23], and then the results of the quantitative data are compared to qualitative data. From the analysis, it has been identified that long term security was preferred by the security engineers, and the financial impact of the information security is difficult to calculate quantitatively. A similar analysis has been performed by [24] on the performance of information security and security awareness of the engineers. Interviews have been performed in an IT company for exploring the personal role of information security workers. It was learned that most of the workers did not have any individual security responsibilities; higher workload increases the conflict between security and performance. They must be involved more into the processes to have a better-balanced improvement in security and performance.

The different types of databases have been discussed in [25] and have focussed on the CryptDB database. The type that provides the most secure among the different types of CryptDB database is Random (RND). This type is attributed to providing security against under a probabilistic adaptive plain text attack. The available data has been split and then mapped into the respective ciphertexts using the vectorized process. Ransom makes sure that the calculations are conducted directly on the ciphertexts. Encryption approaches like AES and Blowfish are performed simultaneously in RND with a random initialization vector. AES is used more than Blowfish for 64-bit block sizes since AES has 128-bit blocks that cause the ciphertexts to be very long.

Researchers of [26] has proposed the vectorization in the form of cryptographic security and digital map in this implementation.

Data encryption has been used to provide security to the network in this work. An iterative encryption organization has been proposed for compressing the vector domain in copy protection and retrieving the control of vector data. The cipher size is reduced by encrypting the various processes through Minimum Coding Object (MCO). The mean points of the data are permutated with position-based encryption, and then the secondary data is fit inside.

A Ciphertext policy with consistent size has been introduced by [27] known as Cipher-text Policy Hidden Vector Encryption (CP-HVE). This type of encryption is unique, where the IBE gets anonymous and utilizes the identity as the main parameters where the attributes linked with the ciphertext or secret keys contain wild-cards. Different arrangements have been utilized, one with a composite ordered bi-linear group while it is a prime ordered bilinear group. These schemes offer higher security by distinguishing the plaintext and ciphertext. This vectorized encryption can have a uniform size than the other HVE methods.

Scholars of [28] has performed quantitative analysis using word2vec to form semantic forms of Naturalistic Driving Data (NDD). Since large amounts of data must be compressed and create compact versions without any loss, the performance has been carried out. Quantitative analysis has been performed to identify the semantic link between the sequences in words. [29] has proposed an automatic intrusion detection system known as MalDozer for the Android operating system. It uses deep learning techniques for classifying the sequences and then word2vec has been used for API for detecting the patterns and malware in the system. The proposed approach has been implemented on servers and evaluated on multiple devices. The result shows the proposed approach and showed high accuracy

According to a study by [30] has also performed quantitative analysis for sentiment analysis. Word2vec method has been used for extracting the features, and then Principal Component Analysis (PCA) was used to identify the important elements and structure. Various classifiers have been compared for performing the work and evaluating the results. In [31] has conducted research for the holistic management of cybersecurity by both quantitative and qualitative methodologies. From the analysis, it was concluded that small and medium-sized businesses had no structural mechanisms to reduce security risks. Instead, they applied the solutions after the breach takes place. Hence, there are not equipped to handle the risks and threats beforehand. The main elements of the framework are identified as external factors, assessing the risks and organizational behaviors.

A study by [32] has proposed a one-shot learning approach for the rapid detection of the attacks. The implementation has been performed for computer vision for recognizing images and texts. Memory Augmented Neural Network has been used in this work, combined with malware detection. Word2vec was used for converting the Application Programming Interface (API) sequences into numeric vectors before getting fed into the one-shot learning network. High accuracy was obtained due to the vectorization using word2vec.

In [33] has proposed a malware detection approach, namely DroidVecDeep, that uses a deep learning approach. Different features are extracted and then ranked by utilizing the Mean Decrease Impurity approach. The features are then transformed into compact vectors on the basis of word2vec. A deep learning model is used to train the classifier, and various malware detection methods are compared. From the results, it is seen that the proposed approach works better than other detection techniques. Hence, vectorization using word2vec is an effective algorithm with deep learning.

A study by [34] has performed group encryption with lattice assumptions through the relation between the matrix vectors. The implementation has been performed through the proposed model, and it was seen that the proposed model provided higher security with certain assumptions. Public key encryption has been used to demonstrate the processes of the novel methodology. However, only generic data has been vectorized rather than the processes.

In [35] has also utilized word2vec for constructing vectors with context sentences and sense the definition vectors giving scores to individual words. WordNet database has been used for retrieving the words and structure and get trained. Scores are based on specific thresholds, and they will be combined with the probability of SEMCOR, which is a sense tagged corpus. It is seen from the results that the performance is better without the probability of distribution. A study by [36] has developed a framework for improving network security. The related studies that focus on network security have been studied with respect to changes to ensure security, passive defenses, the formation of strategy, and overall trend prediction. A detailed critical analysis has been performed by comparing the related articles. The process has included representations, analysis of the solution, predicting the situation, and acquiring the factors.

Encryption is usually performed in such a way that there is very little loss in the vectorized data. A study by [37] has implemented word2vec with Long Short-Term Memory (LSTM) approach for provided security from the attacks. The opcodes have been analyzed in the executable files and maybe clustered using the LSTM network. Word2vec was used for converting the names of the API through one hot encoding results in higher dimensional vectors since the individual cases are represented using the individual dimensions. The proposed approach had better performance than a one-hot encoding-based approach. The issues with data encryption and decryption have been analyzed by [38] through a questionnaire-based methodology. An overview of different schemes for single and multiple secrets of image data has been performed. The different factors that are taken into consideration for the encryption are contrast, capacity, number of shares, type of shares, accuracy, security, and the image format. From these factors, the issues have been identified and have been suggested to address them through novel methods.

Authors of [39] has analyzed the trends in blockchain technology by using word2vec in combination with Latent Semantic Analysis (LSA). Most existing research work has used effort demanding a complete textual investigation and traditional bibliometric approaches. Hence, k means clustering approach has been used for capturing the lexical context. Annual trend analysis has been used for various chains of blockchain. It is seen that the proposed approach has higher accuracy with respect to quantitative methodology. A qualitative approach has also been performed.

From the above review of literature, it can be seen that there are not many studies performing quantitative analysis through

interviews and questionnaires. Most of the studies have directly improved the work by trying to implement the algorithms; however, there are hardly any studies that analyze the management problem by interviewing the public on this. Hence, a comprehensive methodology is required to directly interview the concerned people and collect the data to understand the problem faced and create an effective solution.

3. Methodology

In this work, the quantitative methodology is used to collect the data and analyze the problem. The quantitative methodology can be defined as a type of analysis that uses data that can be easily quantifiable to perform mathematical, computational, and statistical approaches [40]–[52]. Data is collected from the potential sample population through any of the sampling methods in the form of questionnaires [53]. These sets of questions may be asked directly or through the internet in the form of polls or mails. The responses are always converted to the numerical form in such a way that they are quantifiable. The implementation uses interpretivism to understand the problem and gain solutions. Interpretivism is a type of methodology that the researcher implements for synthesizing the facts that are derived from the sources. These facts depend on lots of factors that cannot be measured physically, or they are difficult to measure. The factors may fall under social, cultural, or economic categories. This is analyzed to realize whether the database or security engineers know how to provide security to the database that is vectorized.

The quantitative strategy is applicable for this research since the link between the various variables can be set by interpreting them [54]. This requires interviews to be conducted through questionnaires. The questionnaire addressed the software engineer, who is the most suitable discipline to responded and address the issues in dealing with wrd2vec and CryptDB. Questionnaires are selected for this analysis since they are more reliable and can collect data from lots of respondents quickly and effectively. This is essentially true for larger projects with thousands of respondents. The major disadvantage, however, is that the format is fixed for the entire population and cannot be customized depending on the respondent and this removes the likelihood of in-depth analysis. However, this work does not require variable responses, and hence questionnaires are sufficient. A survey contains various questions relating to the knowledge of vectorization, and security is distributed to the engineers of different companies. They are selected based on the different available companies identified online and their engineers are contacted through email. The responses are captured through online forms and surveys and then compiled together. A total of 30 responses are captured in this way. No sampling strategy was followed here, and only one group of respondents is identified. The questionnaire consists of eleven questions divided into three groups. The groups are dedicated to experience, security, and time. The experience gets from the data related to the educational background and work experience of the respondents. Security aspect relates to the knowledge of the correspondence on security of word2vec and vectorization. Time measures the ability of the security engineers to manage time in performing any changes to the design of the database. The results from the questionnaire are compiled into a single file. Since the number of respondents is very small, detailed analysis is not required. In this work, SPSS software is used for the quantitative analysis.

4. Data Analysis

From the collected results, the analysis performed through SPSS tool is discussed in this section. Table 1 gives details of the qualification of the respondents. A major portion of them is undergraduates at 53.5%, followed by postgraduates at 30%, doctorates at 10%, and certification courses at 6.7%. Table 2 explains the work experience of the respondents. Most of the respondents have experienced between 4 and 8 years, while only 2 respondents had experienced less than four years, suggesting that most of the respondents are well experienced. This is because they are in a senior position. Around 30% of the respondent had more than nine years' experience, out of which 10% had more than 15 years of experience. The respondents who have completed certificate courses have a very high experience, which conveys that only graduates are allowed to become security engineers in the recent decade.

Table 1: Frequency of Qualification

Qualification	Number of respondents	Percent
Computer certification course	2	6.7
Undergraduate	16	53.3
Postgraduate	9	30.0
Doctorate	3	10.0
Total	30	100.0

Table 2: Frequency of Experience

Experience	Number of respondents	Percent
0-3 years	2	6.7
4-8 years	19	63.3
9-15 years	6	20.0
More than 15 years	3	10.0
Total	30	100.0

The ability to solve the security problems in vectorization is given in Table 3. It is seen that most of the respondents are able to solve security issues, with 60% of them pertaining to certain problems, while 33.3% were confident of solving all the problems. Two of the respondents said that they are unable to solve security issues. While more than 90% of the respondents are able to solve security problems, only 66.7% of them have used the CryptDB database for security, as shown in Table 4. The data is also visualized in Figure 2.

Table 3: Ability to solve the security issues during vectorization

Ability to solve the security issues during vectorization	Number of respondents	Percent
Unable to solve the problems	2	6.7
Can solve certain problems	18	60.0
Can effectively solve all the problems	10	33.3
Total	30	100.0

Table 4: Frequency of use of CryptDB database

Use CryptDB database	Number of respondents	Percent
Yes	20	66.7
No	10	33.3
Total	30	100.0

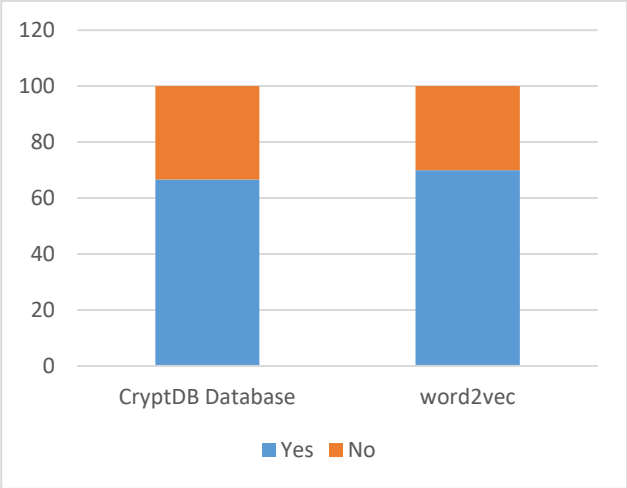

Figure 2: Use of CryptDB and word2vec

A large portion of the respondents has used word2vec embedding for vectorization. Around 70% of them have used them, while the other 30% have either used other types of embedding techniques or never used vectorization as shown in Table 5.

Table 5: Frequency of use of Word2ved embedding

Do you use Word2vec embedding?	Number of respondents	Percent
Yes	21	70.0
No	9	30.0
Total	30	100.0

Since word2vec can be performed easily, it is implemented; however it takes time for embedding them along with the secure database. Only a few of the respondents could complete the implementation within a day, while almost three quarters of the respondents will take from 2 to 7 days. 13.3% of the respondents will take more than a week to complete as shown in Table 6.

Table 6: Time taken for Word2ved embedding

Time is taken for Word2ved embedding	Number of respondents	Percent
Within a day	4	13.3
Takes about a week	22	73.3
Takes more than a week	4	13.3

Once the database is secured, then there might be possibilities of making changes to the design or contents in the work. This might be easy for some minor modifications, where the contents can be modified easily. However, for adding a new component or embed a new element, the design has to be modified, which would take some time. Most of the respondents at around 86.7% of them responded that it would take a long time for them to make changes in the design, only 13.3% of them said it could be done quickly as shown in Table 7.

Table 7: Frequency of changes be made in the design quickly if required

Can any changes be made in the design quickly if required	Number of respondents	Percent
Yes, design can be changed quickly	4	13.3
It takes a long time to make design changes	26	86.7
Total	30	100.0

Even though most respondents have used CryptDB and word2vec, they do not use it in their everyday results. Table 8 shows whether the respondents use CryptDB and word2vec regularly.

Table 8: Frequency of the usage of the concepts

Which one do you use regularly?	Number of respondents	Percent
CryptDB	10	33.3
Word2vec	9	30.0
Both	11	36.7

The mean and the standard deviation for the security is also given in Table 9 for the CryptDB database and word2vec embedding. The p values are seen to be at 0.002, which is considered to be a significant value.

Table 9: Difference in mean security between types of database

	CryptDB database	Word2vec	Both	p-value
	Mean ± SD			
Security	2.73±0.90	2.63±0.51	3.64±0.35	**0.002****

Table 10 gives the opinion of the engineers related to database security. Most of the engineers strongly agree that their database is secure. 40% of the respondents strongly agree, followed by 26.7% agreeing that their database is secure. On the other hand, 32% of the respondents do not agree to this and feel that security must be improved. The majority of the respondents also feel that CryptDB would improve security in the database. 66.7% of the respondents agree, while the other 33.3% disagree that CryptDB improves the security features. While 40% of the respondents strongly agree, only 3.3% of them strongly disagree with CryptDB. Most of the respondents also agree that a secure database must be used for secure vectorizations. 86.7% of the respondents agree that a secure database must be used, while 13.3% of them disagree that security is not required.

The descriptive statistics for the database security are shown in Table 11. Out of the 4-point scale from strongly disagree to strongly agree, a mean of 3.3 is seen along with a standard deviation of 0.76 for the 30 respondents. This is more inclined towards the right with more people agreeing to the conditions in table 10. Since the mean is near 3, more people just agree to the

points while some of them strongly agree. The frequency of disagreeing and strongly disagree is also less.

Table 10: Database security

		N	%
My database is secure	Strongly Disagree	5	16.7
	Disagree	5	16.7
	Agree	8	26.7
	Strongly Agree	12	40.0
CryptDB database improve the security features	Strongly Disagree	1	3.3
	Disagree	9	30.0
	Agree	8	26.7
	Strongly Agree	12	40.0
Secure database must be used for vectorization?	Strongly Disagree	1	3.3
	Disagree	3	10.0
	Agree	16	53.3
	Strongly Agree	10	33.3

Table 11: Descriptive statistics for security

	N	Minimum	Maximum	Mean	SD
Security	30	1.33	4.00	3.03	0.76

5. Conclusions and Recommendations

From the above analysis, it is identified that most of the scientists think that security is necessary for the databases. Even though there are a lot of secure databases and word embedding options, however, around 63% of the interviewed engineers use CryptDB for security purposes, and around 66% use word2vec for embedding options. While many scientists agree that security must be improved in the databases, most scientists completely agree that CryptDB can definitely improve security. Also, most of the engineers accept that word2vec is an effective way to completely improve the vectorization. This work has presented a quantitative methodology for understanding the link between the word vectorization approach word2vec and providing security to the database by using CryptDB. Initially, a brief introduction to both these terminologies concerning vectorization was studied, and then related studies have been studied in the literature review. A questionnaire was created to perform interviews on security engineers. Data is collected in the form of online forms, and then, the data is analyzed through SPSS tool. From the data, it is identified that both word2vec and CryptDB is recommended for embedding the words and security, respectively. The simultaneous use of both of them is suggested as an effective combination for secure vectorization. Hence, the implementation of both of them together is considered as future scope of this work.

6. Limitation and Caveats

The gap identified is that even though many researchers have individually implemented them both. They haven't been implemented together. Also, most of the studies have implemented them technically, but haven't performed interview-based analysis. The limitation of this work was due to the lockdown that happened in March due to COVID-19, and the analysis was conducted on a small scale group of security engineers, the number of correspondents considered in the study was little. In future work, the study will consider both quantitative and qualitative methods. In this study, there are no caveats discussed in this paper.

References

[1] H. Zamani and W. B. Croft, "Relevance-based word embedding," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017, 505–514. https://doi.org/10.1145/3077136.3080831

[2] S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, "Mining in Educational Data: Review and Future Directions," in *Joint European-US Workshop on Applications of Invariance in Computer Vision*, 2020, 92–102. https://doi.org/10.1007/978-3-030-44289-7_9

[3] S. A. Salloum, R. Khan, and K. Shaalan, "A Survey of Semantic Analysis Approaches," in *Joint European-US Workshop on Applications of Invariance in Computer Vision*, 2020, 61–70. https://doi.org/10.1007/978-3-030-44289-7_6

[4] M. Al-Emran, S. Zaza, and K. Shaalan, "Parsing modern standard Arabic using Treebank resources," in *2015 International Conference on Information and Communication Technology Research, ICTRC 2015*, 2015. 10.1109/ICTRC.2015.7156426

[5] C. Mhamdi, M. Al-Emran, and S. A. Salloum, *Text mining and analytics: A case study from news channels posts on Facebook*, vol. 740. 2018. https://doi.org/10.1007/978-3-319-67056-0_19

[6] M. Al-Emran, "Hierarchical Reinforcement Learning: A Survey," *Int. J. Comput. Digit. Syst.*, 4(2), 137–143, 2015. http://dx.doi.org/10.12785/IJCDS/040207

[7] H. Zamani and W. B. Croft, "Estimating embedding vectors for queries," in *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval*, 2016, 123–132. https://doi.org/10.1145/2970398.2970403

[8] O. Press and L. Wolf, "Using the output embedding to improve language models," *arXiv Prepr. arXiv1608.05859*, 2016. https://arxiv.org/abs/1608.05859

[9] K. Hashimoto, C. Xiong, Y. Tsuruoka, and R. Socher, "A joint many-task model: Growing a neural network for multiple nlp tasks," *arXiv Prepr. arXiv1611.01587*, 2016. https://arxiv.org/abs/1611.01587

[10] D. Sarkar, "The Promise of Deep Learning," in *Text Analytics with Python*, Springer, 2019, 631–659. https://doi.org/10.1007/978-1-4842-4354-1_10

[11] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv Prepr. arXiv1301.3781*, 2013. https://arxiv.org/abs/1301.3781

[12] S. Ghosh, M. Chollet, E. Laksana, L.-P. Morency, and S. Scherer, "Affect-lm: A neural language model for customizable affective text generation," *arXiv Prepr. arXiv1704.06851*, 2017. https://arxiv.org/abs/1704.06851

[13] T.-H. Yang, T.-H. Tseng, and C.-P. Chen, "Recurrent neural network-based language models with variation in net topology, language, and granularity," in *2016 International Conference on Asian Language Processing (IALP)*, 2016, 71–74. 10.1109/IALP.2016.7875937

[14] Y. Li and M. Chen, "Software-defined network function virtualization: A survey," *IEEE Access*, 3, 2542–2553, 2015. 10.1109/ACCESS.2015.2499271

[15] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Commun. Surv. tutorials*, 18(1), 236–262, 2015. 10.1109/COMST.2015.2477041

[16] J. F. Kurose, *Computer networking: A top-down approach featuring the internet, 3/E*. Pearson Education India, 2005.

[17] "Intel Corporation (2015). Intel Vectorization Tools. [Online].," 2015.

[18] "Intel (2013). Intel® Xeon PhiTM Processors. [Online]," 2013.

[19] C. Stylianopoulos, L. Johansson, O. Olsson, and M. Almgren, "CLort: High Throughput and Low Energy Network Intrusion Detection on IoT Devices with Embedded GPUs," in *Nordic Conference on Secure IT Systems*, 2018, 187–202. https://doi.org/10.1007/978-3-030-03638-6_12

[20] R. A. Popa, C. M. S. Redfield, N. Zeldovich, and H. Balakrishnan, "CryptDB: processing queries on an encrypted database," *Commun. ACM*, vol. 55, no.

9, 103–111, 2012.
[21] A. Munteanu, "Information security risk assessment: The qualitative versus quantitative dilemma," in *Managing Information in the Digital Economy: Issues & Solutions-Proceedings of the 6th International Business Information Management Association (IBIMA) Conference*, 2006, 227–232. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=917767
[22] S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, "Machine Learning and Deep Learning Techniques for Cybersecurity: A Review," in *Joint European-US Workshop on Applications of Invariance in Computer Vision*, 2020, 50–57. https://doi.org/10.1007/978-3-030-44289-7_5
[23] S. K. Yousuf H., Lahzi M., Salloum S.A., "Systematic Review on Fully Homomorphic Encryption Scheme and Its Application.," *Al-Emran M., Shaalan K., Hassanien A. Recent Adv. Intell. Syst. Smart Appl. Stud. Syst. Decis. Control. vol 295. Springer, Cham*, 2021. https://doi.org/10.1007/978-3-030-47411-9_29
[24] E. Albrechtsen, "A qualitative study of users' view on information security," *Comput. Secur.*, 26(4), 276–289, 2007. https://doi.org/10.1016/j.cose.2006.11.004
[25] R. A. Popa, C. M. S. Redfield, N. Zeldovich, and H. Balakrishnan, "CryptDB: protecting confidentiality with encrypted query processing," in *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles*, 2011, 85–100. https://doi.org/10.1145/2043556.2043566
[26] B.-J. Jang, S.-H. Lee, and K.-R. Kwon, "Perceptual encryption with compression for secure vector map data processing," *Digit. Signal Process.*, vol. 25, 224–243, 2014. https://doi.org/10.1016/j.dsp.2013.09.013
[27] T. V. X. Phuong, G. Yang, and W. Susilo, "Efficient hidden vector encryption with constant-size ciphertext," in *European Symposium on Research in Computer Security*, 2014, 472–487. https://doi.org/10.1007/978-3-319-11203-9_27
[28] Y. Fuchida, T. Taniguchi, T. Takano, T. Mori, K. Takenaka, and T. Bando, "Driving word2vec: Distributed semantic vector representation for symbolized naturalistic driving data," in *2016 IEEE Intelligent Vehicles Symposium (IV)*, 2016, 1313–1320. 10.1109/IVS.2016.7535560
[29] E. B. Karbab, M. Debbabi, A. Derhab, and D. Mouheb, "Android malware detection using deep learning on API method sequences," *arXiv Prepr. arXiv1712.08996*, 2017. https://arxiv.org/abs/1712.08996
[30] X. Ge, X. Jin, and Y. Xu, "Research on sentiment analysis of multiple classifiers based on word2vec," in *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2018, vol. 2, 230–234. 10.1109/IHMSC.2018.10159
[31] J. Jung, "A Study of Cyber Security Management within." University of Portsmouth, 2018.
[32] T. K. Tran, H. Sato, and M. Kubo, "One-shot Learning Approach for Unknown Malware Classification," in *2018 5th Asian Conference on Defense Technology (ACDT)*, 2018, 8–13. 10.1109/ACDT.2018.8593203
[33] T. Chen, Q. Mao, M. Lv, H. Cheng, and Y. Li, "DroidVecDeep: Android Malware Detection Based on Word2Vec and Deep Belief Network.," *TIIS*, 13(4), 2180–2197, 2019.
[34] B. Libert, S. Ling, F. Mouhartem, K. Nguyen, and H. Wang, "Zero-knowledge arguments for matrix–vector relations and lattice-based group encryption," *Theor. Comput. Sci.*, 759, 72–97, 2019. https://doi.org/10.1007/978-3-662-53890-6_4
[35] K. Orkphol and W. Yang, "Word sense disambiguation using cosine similarity collaborates with Word2vec and WordNet," *Futur. Internet*, vol. 11, no. 5, p. 114, 2019. https://doi.org/10.3390/fi11050114
[36] Y. Li, G. Huang, C. Wang, and Y. Li, "Analysis framework of network security situational awareness and comparison of implementation methods," *EURASIP J. Wirel. Commun. Netw.*, 2019(1), 205, 2019. https://doi.org/10.1186/s13638-019-1506-1
[37] J. Kang, S. Jang, S. Li, Y.-S. Jeong, and Y. Sung, "Long short-term memory-based malware classification method for information security," *Comput. Electr. Eng.*, 77, 366–375, 2019. https://doi.org/10.1016/j.compeleceng.2019.06.014
[38] L. R. Logeshwari, R. & Parvathy, "A Quantitative and Qualitative Reasoning of Various Visual Cryptographic Schemes to Uphold Secrecy. International Journal of Innovative Technology and Exploring Engineering. [Online]. 8 (11). 1148–1151.," 2019.
[39] S. Kim, H. Park, and J. Lee, "Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: a study on blockchain technology trend analysis," *Expert Syst. Appl.*, 113401, 2020. https://doi.org/10.1016/j.eswa.2020.113401
[40] C. Q. (2019). Dhir, R.K., de Brito, J., Silva, R. V. & Lye, "Methodology. In: Sustainable Construction Materials.," *[Online]. Elsevier*, 15–34, 2019.
[41] S. A. S. Salloum and K. Shaalan, "Investigating students' acceptance of E-learning system in Higher Educational Environments in the UAE: Applying the Extended Technology Acceptance Model (TAM)." The British University in Dubai, 2018.
[42] M. Habes, M. Alghizzawi, S. Ali, A. SalihAlnaser, and S. A. Salloum, "The Relation among Marketing ads, via Digital Media and mitigate (COVID-19) pandemic in Jordan," *Int. J. Adv. Sci.*, 29(7), 2326–12348, 2020.
[43] M. Habes, M. Alghizzawi, S. A. Salloum, and C. Mhamdi, "Effects of Facebook Personal News Sharing on Building Social Capital in Jordanian Universities," in *Recent Advances in Intelligent Systems and Smart Applications*, Springer, 2020, 653–670. https://doi.org/10.1007/978-3-030-47411-9_35
[44] S. S. A. Al-Maroof R.S., "An Integrated Model of Continuous Intention to Use of Google Classroom.," *Al-Emran M., Shaalan K., Hassanien A. Recent Adv. Intell. Syst. Smart Appl. Stud. Syst. Decis. Control. vol 295. Springer, Cham*, 2021. https://doi.org/10.1007/978-3-030-47411-9_18
[45] S. A. Salloum, A. Q. M. Alhamad, M. Al-Emran, A. A. Monem, and K. Shaalan, "Exploring Students' Acceptance of E-Learning Through the Development of a Comprehensive Technology Acceptance Model," *IEEE Access*, vol. 7, 128445–128462, 2019. 10.1109/ACCESS.2019.2939467
[46] S. A. Salloum and M. Al-Emran, "Factors affecting the adoption of E-payment systems by university students: Extending the TAM with trust," *Int. J. Electron. Bus.*, 14(4), 371–390, 2018. https://doi.org/10.1504/IJEB.2018.098130
[47] S. A. Salloum and K. Shaalan, "Factors affecting students' acceptance of e-learning system in higher education using UTAUT and structural equation modeling approaches," in *International Conference on Advanced Intelligent Systems and Informatics*, 2018, 469–480. https://doi.org/10.1007/978-3-319-99010-1_43
[48] S. A. Salloum and K. Shaalan, "Adoption of e-book for university students," in *International Conference on Advanced Intelligent Systems and Informatics*, 2018, 481–494. https://doi.org/10.1007/978-3-319-99010-1_44
[49] M. Al-Emran, I. Arpaci, and S. A. Salloum, "An empirical examination of continuous intention to use m-learning: An integrated model," *Educ. Inf. Technol.*, 2020. https://doi.org/10.1007/s10639-019-10094-2
[50] S. A. Salloum, M. Al-Emran, R. Khalaf, M. Habes, and K. Shaalan, "An Innovative Study of E-Payment Systems Adoption in Higher Education: Theoretical Constructs and Empirical Analysis," *Int. J. Interact. Mob. Technol.*, 13(6), 2019. https://onlinejour.journals.publicknowledgeproject.org/index.php/i-jim/article/view/9875
[51] R. S. Al-Maroof, S. A. Salloum, A. Q. AlHamadand, and K. Shaalan, "Understanding an Extension Technology Acceptance Model of Google Translation: A Multi-Cultural Study in United Arab Emirates," *Int. J. Interact. Mob. Technol.*, 14(3), 157–178, 2020. https://onlinejour.journals.publicknowledgeproject.org/index.php/i-jim/article/view/11110
[52] M. Habes, S. A. Salloum, M. Alghizzawi, and C. Mhamdi, "The Relation Between Social Media and Students' Academic Performance in Jordan: YouTube Perspective," in *International Conference on Advanced Intelligent Systems and Informatics*, 2019, 382–392.
[53] P. Galbacs, *The Friedman-Lucas Transition in Macroeconomics: A Structuralist Approach*. Academic Press, 2020.
[54] M. Al-Emran, V. Mezhuyev, and A. Kamaludin, "PLS-SEM in Information Systems Research: A Comprehensive Methodological Reference," in *4th International Conference on Advanced Intelligent Systems and Informatics (AISI 2018)*, 2018, 644–653. https://doi.org/10.1007/978-3-319-99010-1_59